

# Selected topics in diffusion Monte Carlo: Lecture III

Claudia Filippi

Instituut-Lorentz, Universiteit Leiden, The Netherlands

Summer School: QMC from Minerals and Materials to Molecules  
July 9-19, 2007, University of Illinois at Urbana-Champaign

Did we remove bias introduced by wave function?

We wanted to go beyond VMC → Fixed-node approximation

Results depend on the nodes of the trial wave function  $\Psi$

How well does it work if we do not worry too much about  $\Psi$ ?

## DMC as black-box approach?

How well are we doing with a simple  $\Psi$ ?

One determinant of natural orbitals, 6-311++G(2d,2p) basis

Atomization energy of the 55 molecules of the G1 set

Mean absolute deviation from experiments  $\epsilon_{\text{MAD}}$

	QMC	CCSD(T)/aug-cc-pVQZ
$\epsilon_{\text{MAD}}$	2.9	2.8 kcal/mol

Grossman, J. Chem. Phys. **117**, 1434 (2002)

We are doing very well without much effort on  $\Psi$  !

How far can we go with the no-brain-no-pain approach?

	QMC	CCSD(T)	DFT/B3LYP
$\epsilon_{\text{MAD}}$	2.9	2.8	2.5 kcal/mol

Disappointing how well B3LYP works!

But with some more effort ... QMC can do much better!

Example: Atomization energy of  $\text{P}_2$

DMC one-det	107.9(2)	kcal/mol
DMC multi-det	115.9(2)	
Experiment	116.1(5)	

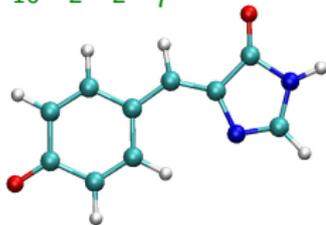
Grossman, J. Chem. Phys. **117**, 1434 (2002)

## Alleviating wave function bias by optimization

How do we obtain the parameters in the wave function?

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J} \sum_k d_k D_k^\uparrow D_k^\downarrow$$

$\text{C}_{10}\text{N}_2\text{O}_2\text{H}_7^-$



70 electrons and 21 atoms

VTZ *s-p* basis + 1 polarization

3 *s* + 3 *p* + 1 *d* functions for C, N, O

2 *s* + 1 *p* for H

- ▷ Parameters in the Jastrow factor  $\mathcal{J}$  ( $\approx 100$ )
- ▷ CI coefficients  $d_k$  ( $< 10$ )
- ▷ Linear coefficients in expansion of the orbitals (5540 !)

## Customary practice for optimizing wave function

### Jastrow-Slater wave function

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J} \sum_k d_k D_k^\uparrow D_k^\downarrow$$

- ▷ Jastrow factor optimized in variance/energy minimization
- ▷ Orbitals and  $d_k$  coefficients in determinantal part are from
  - Hartree-Fock or DFT (LDA, GGA, B3LYP ...)
  - CI or multi-configuration self-consistent-field calculation
  - Optimized in variance minimization (small systems)
  - Optimized in energy minimization (very simple for  $d_k$ )

## Optimization of trial wave function

How do we find the **best** parameters in  $\Psi = \mathcal{J}\Phi$  ?

**First thought** Let us minimize the energy!

$$E_V = \frac{\int d\mathbf{R} \Psi^*(\mathbf{R}) \mathcal{H} \Psi(\mathbf{R})}{\int d\mathbf{R} \Psi^*(\mathbf{R}) \Psi(\mathbf{R})} = \int d\mathbf{R} \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} = \left\langle \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \right\rangle_{\Psi^2}$$

Straightforward minimization on finite MC sample will **not** work!

## Why problems with straightforward energy minimization ?

Let us write the energy on a finite MC sample

Sample  $N_{\text{conf}}$  configurations from  $|\Psi(\mathbf{R}, \{\alpha_0\})|^2$  with Metropolis

Energy of  $\Psi(\mathbf{R}, \{\alpha\})$  on this set of MC configurations

$$E[\alpha] = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} w_i$$

where

$$w_i = \left| \frac{\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha_0\})} \right|^2 \bigg/ \sum_{i=1}^{N_{\text{conf}}} \left| \frac{\Psi(\mathbf{R}, \{\alpha\})}{\Psi(\mathbf{R}, \{\alpha_0\})} \right|^2$$

$E[\alpha]$  on a finite MC sample is not bounded from below

$\Rightarrow$  Straightforward minimization of  $E[\alpha]$  does not work

## Is variance minimization an alternative?

Minimize the variance of the local energy

$$\sigma^2 = \frac{\langle \Psi | (\mathcal{H} - E_V)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle (E_L(\mathbf{R}) - E_V)^2 \rangle_{\Psi^2}$$

Would this work?

Consider variance on a finite number of MC configurations

$$\sigma^2[\alpha] = \sum_{i=1}^{N_{\text{conf}}} \left( \frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} - \bar{E} \right)^2 w_i$$

$\sigma^2$  has a known lower bound  $\sigma^2 = 0$

Robust and stable optimization for very small values of  $N_{\text{conf}}$

## Variance minimization

(1)

- ▷ Variance minimization on a fixed set of MC configurations

$$\sigma^2[\alpha] = \sum_{i=1}^{N_{\text{conf}}} \left( \frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} - \bar{E} \right)^2 w_i$$

- $\bar{E}$  substituted with  $E_{\text{guess}}$  a bit smaller than current estimate  $\bar{E}$ 
    - ⇔ Minimize a combination of variance and energy
  - $w_i$  must be limited to a max value (or some  $R_i$  may dominate)
  - $N_{\text{conf}} = 2000-3000$  for 100 parameters in as many as 800 dim
- ▷ Variance minimization on-the-fly by computing gradient/Hessian
- See next week lecture by Umrigar

## Variance minimization

(2)

### Other advantages

- ▷ All eigenstates have zero variance
  - ⇒ It is possible to optimize true excited states
- ▷ Cusp conditions or other constraints easily added
  - ⇒ Minimize  $\chi^2 = \sigma^2 + \text{penalty functions}$
- ▷ Efficient procedures to optimize a sum of squares  
e.g. Levenberg-Marquard

### Main disadvantage

- ▷ It is variance not energy minimization

What about energy minimization?

We want the parameters in  $\Psi$  which give lowest VMC energy

But it would seem simple !?!

Let us compute gradient and Hessian of the energy in VMC

## Energy minimization and statistical fluctuations

Wave function  $\Psi$  depends on parameters  $\{\alpha_k\}$

Energy and derivatives of the energy wrt parameters  $\{\alpha_k\}$  are

$$E_V = \int d\mathbf{R} \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} = \langle E_L \rangle_{\Psi^2}$$
$$\partial_k E_V = \left\langle \frac{\partial_k \Psi}{\Psi} E_L + \frac{\mathcal{H} \partial_k \Psi}{\Psi} - 2E_V \frac{\partial_k \Psi}{\Psi} \right\rangle_{\Psi^2}$$
$$= 2 \left\langle \frac{\partial_k \Psi}{\Psi} (E_L - E_V) \right\rangle_{\Psi^2}$$

The last expression is obtained using Hermiticity of  $\mathcal{H}$

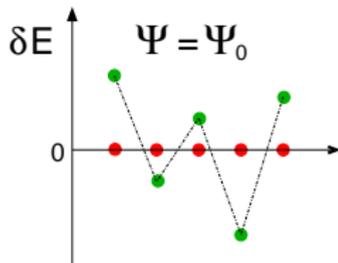
Use gradient/Hessian expressions with smaller fluctuations

Two mathematically equivalent expressions of the energy gradient

$$\partial_k E_V = \left\langle \frac{\partial_k \Psi}{\Psi} E_L + \frac{\mathcal{H} \partial_k \Psi}{\Psi} - 2E_V \frac{\partial_k \Psi}{\Psi} \right\rangle_{\Psi^2} = 2 \left\langle \frac{\partial_k \Psi}{\Psi} (E_L - E_V) \right\rangle_{\Psi^2}$$

Why using the last expression?

Lower fluctuations  $\rightarrow 0$  as  $\Psi \rightarrow \Psi_0$

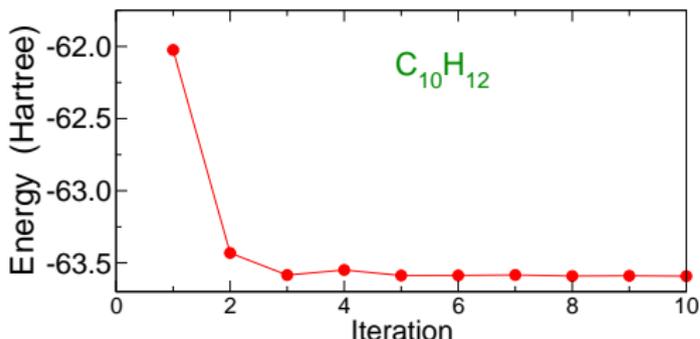


Computation of Hessian  $\rightarrow$  Play similar tricks as for the gradient!

Rewrite expression in terms of covariances

$\rightarrow \langle ab \rangle - \langle a \rangle \langle b \rangle$  usually smaller fluctuations than  $\langle ab \rangle$

Optimization of  $\mathcal{J}$



5 orders of magnitude efficiency gain wrt use of original Hessian

C. Umrigar and C. Filippi, PRL **94**, 150201 (2005)

## Energy minimization is possible

... with simple modifications of straightforward approach

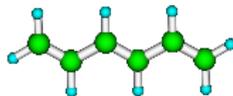
Various energy minimization schemes are available:

- ▷ Stochastic reconfiguration (Sorella, Casula)
- ▷ Energy fluctuation potential (Fahy, Filippi, Prendergast, Schautz)
- ▷ Perturbative method (Filippi, Scemama)
- ▷ Hessian method (Umrigar, Filippi, Sorella)
- ▷ Linear method (Nightingale, Umrigar, Toulouse, Filippi, Sorella)

See next week lecture by Umrigar

## Importance of optimizing the wave function

Example: Excitation energy of hexatriene ( $C_6H_8$ )



State	Wave function	$E_{VMC}$	$E_{DMC}$	$\Delta E$ (eV)
$1^1A_g$	HF	-38.684(1)	-38.7979(7)	—
	B3LYP	-38.691(1)	-38.7997(7)	—
	optimized	-38.691(1)	-38.7992(7)	—
$1^1B_u$	CAS(2,2)	-38.472(1)	-38.5910(7)	5.63(3)
	B3LYP	-38.482(1)	-38.6030(7)	5.35(3)
	optimized	-38.493(1)	-38.6069(8)	5.23(3)
expt.			5.22	

## Fixed-node diffusion Monte Carlo and excited states

Finds the best solution with the same nodes as trial  $\Psi$

Is fixed-node DMC variational?

For lowest state in each 1-dim irreducible representation

## What about “real” excited states?

In general, exact excited state for exact nodal structure

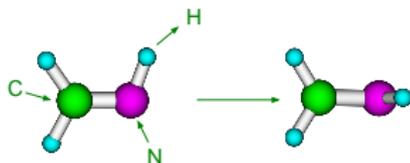
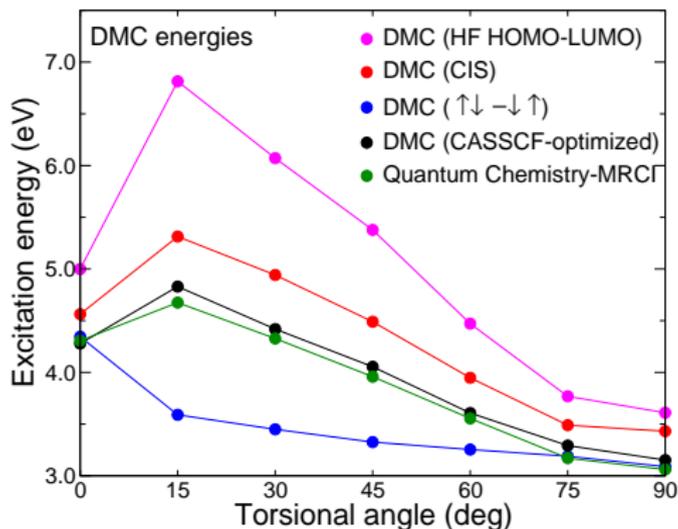
For excited states, even bigger role of the trial wave function

→ Enforces fermionic antisymmetry + selects the state

## Excited states and the trial wave function

Dependence of DMC energy from wave function  $\Psi = \mathcal{J} \left[ \sum_i c_i D_i \right]$

Lowest singlet excitation along torsional path of formaldimine



At  $0^\circ$  and  $90^\circ$ , ground and excited states have different symmetry

Otherwise, same symmetry  $\rightarrow$  "Real" excited state

## Excited state optimal wave function

Wave functions for multiple states of the same symmetry

$$\Psi_I(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_i c_i^I \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) \times D_i(\mathbf{r}_1, \dots, \mathbf{r}_N)$$

Common set of parameters in  $\mathcal{J}$  and  $D_i$  but different coefficients  $c_i^I$

Optimize parameters in  $\mathcal{J}$  and  $D_i$  by state averaging

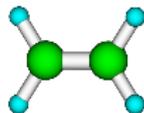
$$E_{\text{SA}} = \sum_I w_I \frac{\langle \Psi_I | \mathcal{H} | \Psi_I \rangle}{\langle \Psi_I | \Psi_I \rangle}$$

and preserve orthogonality through coefficients  $c_i^I$

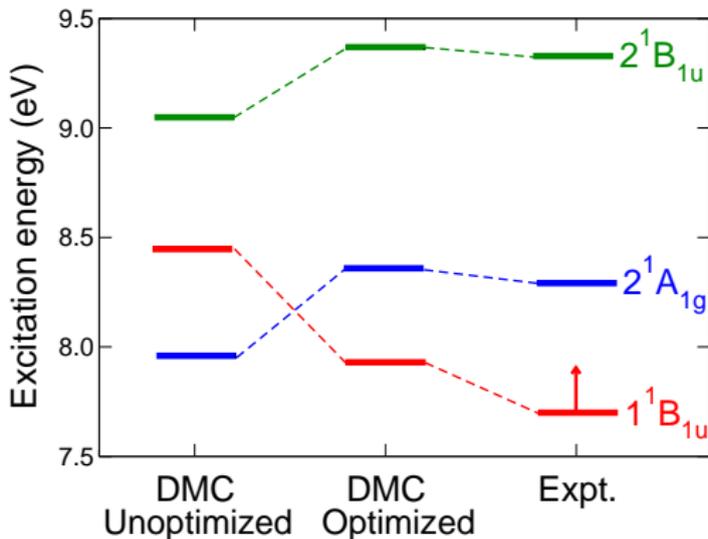
Schautz and Filippi, J. Chem. Phys. **120**, 10931 (2004); Filippi (2007)

## Excitation energies of ethene $C_2H_4$

Difficulties: Valence-Rydberg mixing + core relaxation



Schautz and Filippi, JCP (2004)



## Correlated sampling in VMC

Given two operators  $\mathcal{O}$ ,  $\mathcal{O}'$  and wave functions  $\Psi$ ,  $\Psi'$ , compute

$$\bar{\mathcal{O}}' - \bar{\mathcal{O}} = \frac{\langle \Psi' | \mathcal{O}' | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle} - \frac{\langle \Psi | \mathcal{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

Do NOT perform independent runs  $\rightarrow$  Use **correlated sampling**

To compute differences more accurately than separate quantities

Example: Map out potential energy surface  $\Rightarrow$  Compute  $\Delta E$

DFT/QC methods  $\rightarrow$  Smoothly varying (systematic) error

Again problems in QMC with statistical fluctuations!

## Interatomic forces and geometry optimization

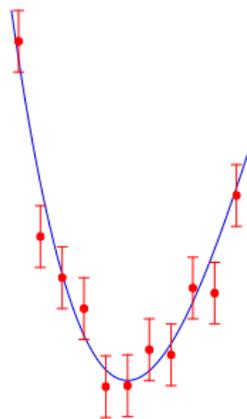
Customary practice: Use DFT or QC geometries

One possible route → Compute forces by finite differences

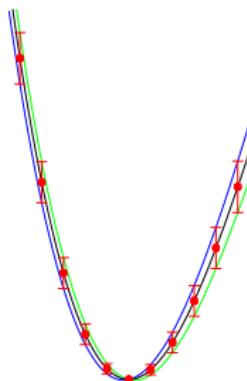
Why problems with statistical fluctuations?

Example: Energy of a dimer versus bond length

Independent runs



Correlated sampling



⇒ Forces cannot be computed from independent runs

## Correlated sampling: The computation of potential energy surfaces

Primary geometry  $\rightarrow \mathcal{H} \quad \Psi \quad E$

Secondary geometry  $\rightarrow \mathcal{H}_s \quad \Psi_s \quad E_s$

$$E_s - E = \frac{\langle \Psi_s | \mathcal{H}_s | \Psi_s \rangle}{\langle \Psi_s | \Psi_s \rangle} - \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

No independent runs  $\rightarrow$  MC configurations only from reference  $\Psi^2$

$$E_s - E = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \left\{ \frac{\mathcal{H}_s \Psi_s(\mathbf{R}_i)}{\Psi_s(\mathbf{R}_i)} w_i - \frac{\mathcal{H} \Psi(\mathbf{R}_i)}{\Psi(\mathbf{R}_i)} \right\}$$

$$w_i = \frac{|\Psi_s(\mathbf{R}_i)/\Psi(\mathbf{R}_i)|^2}{\frac{1}{N_{\text{conf}}} \sum_{j=1}^{N_{\text{conf}}} |\Psi_s(\mathbf{R}_j)/\Psi(\mathbf{R}_j)|^2}$$

Efficient if  $w_i \approx 1$  and  $\mathcal{H}$  and  $\mathcal{H}_s$  closely related

## Efficiency gain from correlated sampling

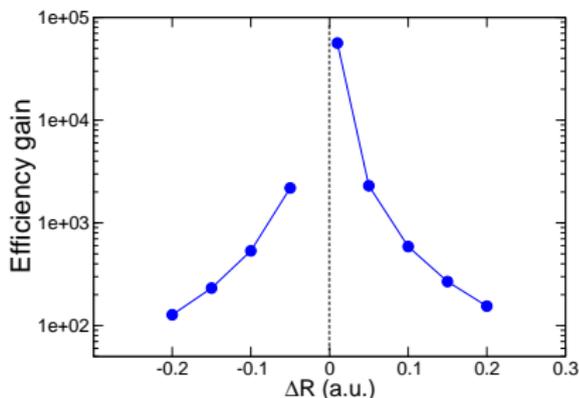
Example:  $B_2$ , 1 determinant + simple Jastrow factor

$E$  at experimental equilibrium bond length  $R_0 = 3.005$  a.u.

$E_s$  at stretched bond length by  $\Delta R = -0.2, \dots, 0.2$  a.u.

Compute  $E_s - E$  from independent runs  $\rightarrow \Delta E_{\text{ind}}$   
from correlated sampling  $\rightarrow \Delta E_{\text{corr}}$

$$\text{Efficiency gain} = \frac{\sigma^2(\Delta E_{\text{ind}})}{\sigma^2(\Delta E_{\text{corr}})}$$



Note: We used space-warp coordinate transformation



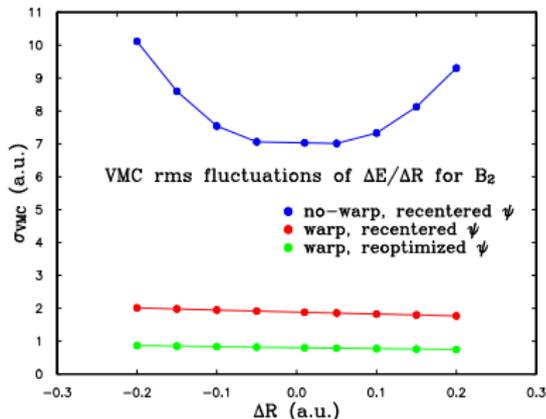
# Energy difference with space-warp transformation

$$E_S - E = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \left\{ \frac{\mathcal{H}_S \Psi_S(\mathbf{R}_i^S)}{\Psi_S(\mathbf{R}_i^S)} w_i - \frac{\mathcal{H} \Psi(\mathbf{R}_i)}{\Psi(\mathbf{R}_i)} \right\}$$

$$\text{with } w_i = \frac{|\Psi_S(\mathbf{R}_i^S)/\Psi(\mathbf{R}_i)|^2 \boxed{J(\mathbf{R}_i)}}{\frac{1}{N_{\text{conf}}} \sum_{j=1}^{N_{\text{conf}}} |\Psi_S(\mathbf{R}_j^S)/\Psi(\mathbf{R}_j)|^2 \boxed{J(\mathbf{R}_j)}}$$

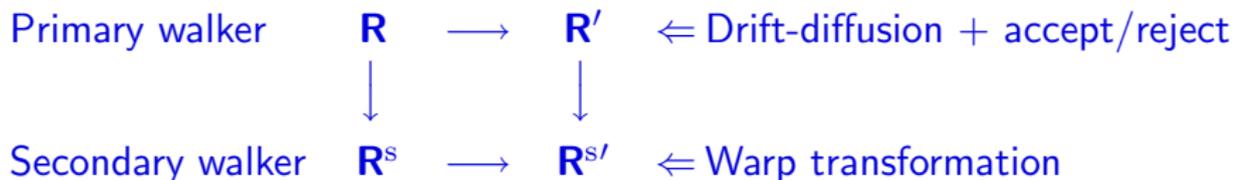
and  $J(\mathbf{R})$  Jacobian of transformation  $\mathbf{R} \rightarrow \mathbf{R}^S$

rms fluctuations of  $F = \frac{\Delta E}{\Delta R}$





## Correlated sampling in DMC



## PROBLEMS

- ▷ Dynamics of secondary walker is wrong
  - $\mathbf{R} \longrightarrow \mathbf{R}'$  with  $T(\mathbf{R}', \mathbf{R}, \tau)$
  - $\mathbf{R} \longrightarrow \mathbf{R}^{s'}$  with warp transformation
    - $\Rightarrow$  Transition with  $T(\mathbf{R}', \mathbf{R}, \tau)/J(\mathbf{R})$  **NOT**  $T_s(\mathbf{R}^{s'}, \mathbf{R}^s, \tau)$
    - $\Rightarrow$  Secondary move accepted with probability  $p$  **NOT**  $p_s$
- ▷ Different nodes for primary and secondary walker

Correlated sampling in DMC: Approximate but accurate scheme

Observation: Correlated sampling in VMC is very efficient

⇒ Scheme similar to VMC but with results very close to DMC?

Filippi, Umrigar, Phys. Rev. B **61**, R16291 (2000)



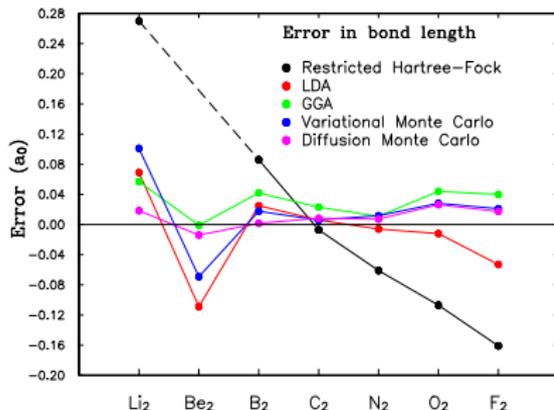
## Correlated sampling: Error in bond length of 1st-row dimers

VMC and DMC from the PES obtained by correlated sampling

RMS errors of bond length (a.u.)

RHF	LDA	GGA	VMC	DMC
$\infty$	0.054	0.036	0.049	0.014

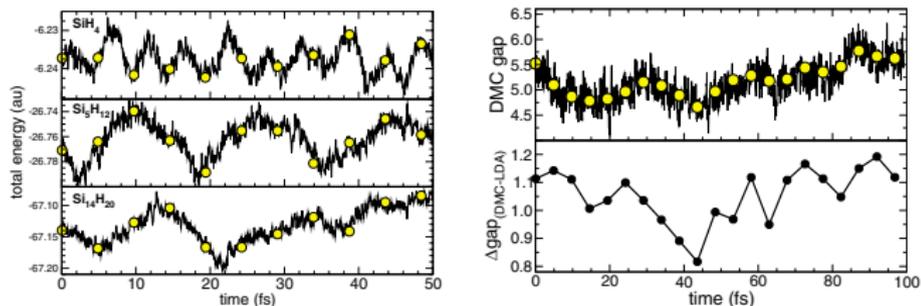
DMC always improves upon VMC



# Continuum Diffusion Monte Carlo

Grossman and Mitas, Phys. Rev. Lett. **94**, 056403 (2005)

Efficient on-the-fly computation of DMC energies on AIMD path



- ▷ At  $t_{\text{MD}} = 0$ ,  $\mathcal{H}(0)$ . Equilibrate DMC population  $\{\mathbf{R}, w\}$
- ▷ At  $t_{\text{MD}} = 1$ ,  $\mathcal{H}(1)$ . Start from previous  $\{\mathbf{R}\}$  and adjust  $w$  as

$$w = w \left| \frac{\Psi(\mathbf{R}; \mathcal{H}(1))}{\Psi(\mathbf{R}; \mathcal{H}(0))} \right|^2 \frac{\exp[-E_L(\mathbf{R}; \mathcal{H}(1))\tau]}{\exp[-E_L(\mathbf{R}; \mathcal{H}(0))\tau]}$$

- ▷ Few DMC steps (3!) sufficient to equilibrate population  $\{\mathbf{R}', w'\}$

What about Hellman-Feynman theorem?

Consider  $\mathcal{H}(\lambda)$  with  $\lambda$  parameter (nuclear coordinates)

$$E(\lambda) = \frac{\langle \Psi(\lambda) | \mathcal{H}(\lambda) | \Psi(\lambda) \rangle}{\langle \Psi(\lambda) | \Psi(\lambda) \rangle} \quad \text{and} \quad \frac{dE(\lambda)}{d\lambda} = \frac{\langle \Psi(\lambda) | \frac{d\mathcal{H}(\lambda)}{d\lambda} | \Psi(\lambda) \rangle}{\langle \Psi(\lambda) | \Psi(\lambda) \rangle}$$

True if  $\Psi(\lambda)$  is an eigenstate or  $\Psi_\alpha(\lambda)$  minimizes energy wrt  $\alpha$

Problems with Hellman-Feynman forces in QMC

## Hellman-Feynman forces in QMC

- Large fluctuations  $\rightarrow$  Infinite for all-electron calculations!

$$\partial_{\alpha} E = \langle \partial_{R_{\alpha}} \mathcal{H} \rangle_{\Psi^2} = \langle F_{\alpha} \rangle_{\Psi^2} \approx \langle \frac{1}{r^2} \rangle_{\Psi^2} = \text{finite}$$

$$\sigma^2(F_{\alpha}) = \langle F_{\alpha}^2 \rangle_{\Psi^2} - \langle F_{\alpha} \rangle_{\Psi^2}^2 = \infty$$

Solution: Reduced variance method by Assaraf-Caffarel

$$\tilde{F}_{\alpha} = F_{\alpha} + \Delta F_{\alpha} \text{ with } \langle \Delta F_{\alpha}^2 \rangle_{\Psi^2} = 0 \text{ but } \sigma^2(\tilde{F}_{\alpha}) \text{ finite}$$

- If  $\Psi$  does not minimize  $E_{\text{VMC}} \Rightarrow$  Systematic error in VMC  
Use energy-minimized wave functions

Application Ab-initio MD for high-pressure liquid Hydrogen  
Sorella and Attaccalite, cond-mat/0703800

Computation of forces/MD: Active field of research in QMC

## Human and computational cost of a typical QMC calculation

Task	Human time	Computer time
Choice of basis set, pseudo etc.	10%	5%
DFT/HF/CI runs for $\Psi$ setup	65%	10%
Optimization of $\Psi$	20%	30%
DMC calculation	5%	55%

- ▶ General reviews:  
Foulkes *et al.*, Rev. Mod. Phys. **73**, 33 (2001);  
Several articles in “Quantum Monte Carlo methods in Physics and Chemistry”, edited by Nightingale and Umrigar (Nato Science Series, 1999).
- ▶ Early applications of variational Monte Carlo:  
McMillan, Phys. Rev. **138**, A4422 (1965);  
Ceperley, Chester, Kalos, Phys. Rev. B **16**, 3081 (1977).
- ▶ VMC and choice of transition matrix:  
Umrigar, Phys. Rev. Lett. **71**, 408 (1993);  
Umrigar in “Quantum Monte Carlo methods in Physics and Chemistry”, edited by M.P. Nightingale and C.J. Umrigar (Nato Science Series, 1999).

## References

(2)

- ▶ Cusp conditions:  
Pack and Byers Brown, J. Chem. Phys. **45**, 556 (1966).
- ▶ Spin-assigned wave function, spin contamination:  
Huang, Filippi, Umrigar, J. Chem. Phys. **108**, 8838 (1998).
- ▶ Jastrow factor and role of e-e-e-n terms:  
Huang, Umrigar, Nightingale, J. Chem. Phys. **107**, 3007 (1997).
- ▶ Some references on diffusion Monte Carlo:  
Reynolds *et al.*, J. Chem. Phys. **77** 5593 (1982);  
Mitas, Shirley, and Ceperley, J. Chem. Phys. **95**, 3467 (1991);  
Umrigar, Nightingale, and Runge, J. Chem. Phys. **99**, 2865 (1993).

- ▶ Variance minimization:  
Coldwell, Int. J. Quantum Chem. Symp. **11**, 215 (1977).  
Umrigar, Wilson, Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).  
Filippi and Umrigar, J. Chem. Phys. **105**, 213 (1996).
- ▶ Correlated sampling for interatomic forces:  
Filippi and Umrigar, Phys. Rev. B **61**, R16291 (2000).
- ▶ Zero-variance, zero-bias principle and Hellman-Feynman forces:  
Assaraf and Caffarel, Phys. Rev. Lett. **83**, 4682 (1999);  
Assaraf and Caffarel, J. Chem. Phys. **113**, 4028 (2000);  
Assaraf and Caffarel, J. Chem. Phys. **119**, 10536 (2003);  
Sorella and Attaccalite, cond-mat/0703800